

# TrajAnalytics: A Web-Based Visual Analytics Software of Urban Trajectory Data

Ye Zhao                      Shamal Al-Dohuki                      Thomas Eynon                      Farah Kamw                      David Sheets  
Kent State University      Kent State University      Kent State University      Kent State University      Kent State University

Chao Ma                      Yueqi Hu                      Xinyue Ye                      Jing Yang  
Kent State University      UNC Charlotte                      Kent State University      UNC Charlotte

## ABSTRACT

We present a web-based software, named TrajAnalytics, for the visual analytics of urban trajectory datasets. It allows users to interactively visualize and analyze the massive taxi trajectories over urban spaces. The software offers data management capability and enable various visual queries through a web interface. A set of visualization widgets and interaction functions are developed to promote easy user engagement. We implemented two prototypes with real-world datasets: (1) the origin/destination (OD) data of taxi trips collected in New York City; and (2) the taxi trajectory data collected in Porto City in Portugal. This open source software supports practitioners, researchers, and decision-makers in advancing transportation and urban studies in the new era of smart city.

**Keywords:** Visual Analytics Software, Taxi Trajectory, Urban Transportation

## 1 INTRODUCTION

Advanced technologies in sensing and computing have created urban trajectory datasets of humans and vehicles travelling over urban networks (e.g., roads and public transits). Understanding and analyzing the large-scale, complex data reflecting city dynamics is a great importance to enhance both human lives and urban environments. Domain practitioners, researchers, and decision-makers need to store, manage, query and visualize such big, dynamic data to conduct exploratory and analytical tasks. However, they are facing great challenges due to the lack of available software supporting data driven study. The software should integrate scalable trajectory databases with intuitive and interactive visualizations, and also facilitate easy access. The absence of such open source tools that explicitly support interactive visual analytics impedes the advancement in the utilization of the emerging trajectory data. In order to support general users, we have developed an open source software system named TrajAnalytics. It offers data management capability and support various data queries by leveraging web-based computing platforms. It allows users to visually conduct queries and make sense of massive trajectory data. The system is made publicly accessible at <http://vis.cs.kent.edu/software.html> under the BSD license.

### 1.1 Urban Trajectory Data

Nowadays, a large amount of trajectory data sets are collected by transportation administrations, companies, and researchers. Each trajectory data records realtime moving paths sampled as a sequence of time-stamped GPS points over urban networks. Rich and heterogeneous information can be associated with each point including time stamp, geographical features such as latitude and longitude, human and vehicle attributes, business/urban information,

and more. Such data is big, spatial, temporal, dynamic, and unstructured. For example, the total length of the 3-month trajectories over 33,000 Beijing taxis [11] is more than 400 million kilometers and the total number of GPS points reaches 790 million. The associated data of the point samples further increases the data scale.

We use two taxi trajectory datasets in developing TrajAnalytics. Other types of urban trajectories will be incorporated later. To make the software publicly accessible, two public datasets are utilized in our prototypes. The first one is an O/D (origin/destination) dataset from the taxi trips in New York city [8]. Each taxi trip consists of car ID, driver ID, pickup time, dropoff time, passenger count, trip time and distance, and pickup and dropoff location (longitude, latitude). The second dataset is the whole taxi trajectory data of Porto city, Portugal [7].

### 1.2 Visual Analytics Software Design

Iterative visual exploration is one key component in processing such data, which should be supported by efficient data management and visualization tools. Therefore, transportation researchers demand a handy and effective visual analytics software system which integrates scalable data management and interactive visualization with powerful computational capability. An ideal visual analytics software system should provide:

- Powerful computing platform so that domain users are not limited by their computational resources and can complete their tasks over daily-used computers or mobile devices.
- Easy access gateway so that the trajectory data can be retrieved, analyzed and visualized by different transportation researchers, and their results can be shared and leveraged by others.
- Scalable data storage and management which support a variety of data queries with immediate responses.
- Exploratory visualizations that are informative, intuitive, and facilitate efficient interactions.
- A multi-user system which allows simultaneous operations by many users from different places.

### 1.3 Related Work

Conventional transportation design software, such as TransCAD [2], Cube [4], provides platforms for urban transportation forecasting, planning and analysis. Domain researchers can build transport models, perform simulations, and create simple visual representations of the results for analysis and presentation. However these software packages are not developed for data driven analysis utilizing realworld trajectory data. Urban computing has emerged recently in the data mining community to advance discovery of urban knowledge from a variety of data including trajectories. However, there is no visual analytics software devoted for domain researchers to utilizing the data [12]. General-purpose information visualization software (e.g. Tableau) partly support the study of trajectory data. Specifically, a set of recent visual analytics approaches have contributed many technologies with prototypes based on the trajectory datasets (e.g., [1, 6, 9, 5, 10]). Please see a survey for more information [3]. However, these approaches are not

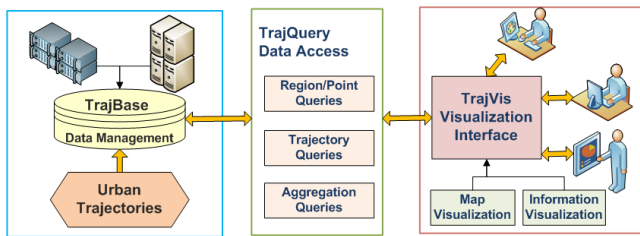


Figure 1: TrajAnalytics Software Framework.

developed to provide publicly accessible and easy-to-use software for general domain researchers. They do not easily satisfy all the above requirements of an open source software.

TrajAnalytics is developed to fill the need for urban transportation researchers in studying and utilizing big trajectory datasets. It will be shared to domain users as a freely accessible software on <http://vis.cs.kent.edu/software.html>. Its aim is to advance the research activities and investigative studies of a wide community in transportation study.

## 2 SYSTEM FRAMEWORK

TrajAnalytics consists of three components: scalable data management (TrajBase), effective data query (TrajQuery), and interactive visual interface (TrajVis). Figure 1 shows the software framework.

### 2.1 TrajBase

A scalable database is specifically designed for storing and managing big trajectory data. We have tested both MySQL, PostGreSQL, and MongoDB databases to fulfill the requirement. The special extensions of these databases of spatial data indexing are utilized. In the prototypes, MySQL is used and B-tree indexing is employed to optimize the data queries. TrajBase supports fast computation over various data queries in a remote and distributed computing environment.

### 2.2 TrajQuery

TrajQuery supports the user to conduct spatial queries combined with temporal constraints. The spatial queries extract taxi trips or trajectories with (1) pick-up regions; (2) drop-off regions; and (3) traversed (passed) regions. A map-based interface is provided to define the regions, including rectangular, circular and free form selections. The three types can be conducted individually or jointly. They are also integrated with time period selections.

### 2.3 TrajVis

To support online visual analysis with fast speed and easy user interaction, we design the system with a set of coordinated views. The interface is illustrated in Figure 2 with the map view, list view of queries, side-by-side view, and visual report view. In implementation, we use the open source libraries of D3.js for information visualization and leaflet.js for map-based visualization. Users can conduct exploratory visual analysis through the informative and intuitive interactions. In particular, users can choose different types of maps for geographic contexts. They can conduct region selections and time period selections through easy mouse operations. The queried results are shown in points, heat maps, or trajectories while users can alter the display methods. To make analysis easier, the list view of queries help users manage multiple queries. Moreover, we design the side-by-side view (Figure 3 left) so that users can conveniently compare different query results. The visual report view (Figure 3 right) shows a set of charts and diagrams of the query results for quantitative attribute analysis. The online visual system also support multiple users to conduct their work from different sites.

## 3 PROTOTYPES AND USER FEEDBACK

Two prototypes of New York (O/D) and Porto (full trajectory) are implemented and published online for test use. We also made two tutorial videos to demonstrate how to use the software with a complete description of functions. Please visit <http://vis.cs.kent.edu/software.html> to access them. We have contacted multiple types of domain users to use the prototypes and provide feedback and suggestions to improve our system.

## 4 FUTURE WORK

We have completed the first stage of software development. We are working on next-stage software development in multiple directions including: new data model and databases for handling extremely large datasets; data aggregation methods in pre-computation to enable fast query performance; cloud computing tools and algorithms; and more visualization widgets and interaction functions. We will particularly design visualization and analytical functions to fulfill specific requirements for users from different fields. We will recruit and train domain students for testing the functional modules of software, and perform complete software evaluation.

## 5 CONCLUSION

The mobility and behavior of moving humans and transportation vehicles form the basic component in human society. Our software facilitates easy, online exploration of big trajectory data. It will advance a broad spectrum of applications by enabling researchers to visually analyze the emerging trajectory data. In addition to that we will work with our collaborators from geographic department to support our work with real-world examples.

## ACKNOWLEDGEMENTS

This work is partially supported by U.S. National Science Foundation under grant numbers 1535031, 1535081, and 1416509.

## REFERENCES

- [1] N. Andrienko and G. Andrienko. Research homepage. Website. <http://geoanalytics.net/and/>.
- [2] Caliper. TransCAD Transportation Planning Software. Website, 2014. <http://www.caliper.com/tcovu.htm>.
- [3] W. Chen, F. Guo, and F.-Y. Wang. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2970–2984, 2015.
- [4] Citilab. Cube. Website, 2014. <http://www.citilabs.com/>.
- [5] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, Dec. 2013.
- [6] H. Liu, Y. Gao, L. Lu, S. Liu, L. Ni, and H. Qu. Visual analysis of route diversity. *IEEE Conference on VAST*, pages 171–180, 2011.
- [7] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas. Predicting taxi passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):393–1402, 2013.
- [8] New York City Taxi Data. <http://publish.illinois.edu/dbwork/open-data/>.
- [9] J. Pu, S. Liu, Y. Ding, H. Qu, and L. Ni. T-watcher: A new visual analytic system for effective traffic surveillance. In *Proceedings of Mobile Data Management*, pages 127–136, 2013.
- [10] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168, 2013.
- [11] J. Yuan, Y. Zheng, X. Xie, and G. Sun. T-drive: Enhancing driving directions with taxi drivers’ intelligence. *IEEE Trans. on Knowl. and Data Eng.*, 25(1):220–232, Jan. 2013.
- [12] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 2014.

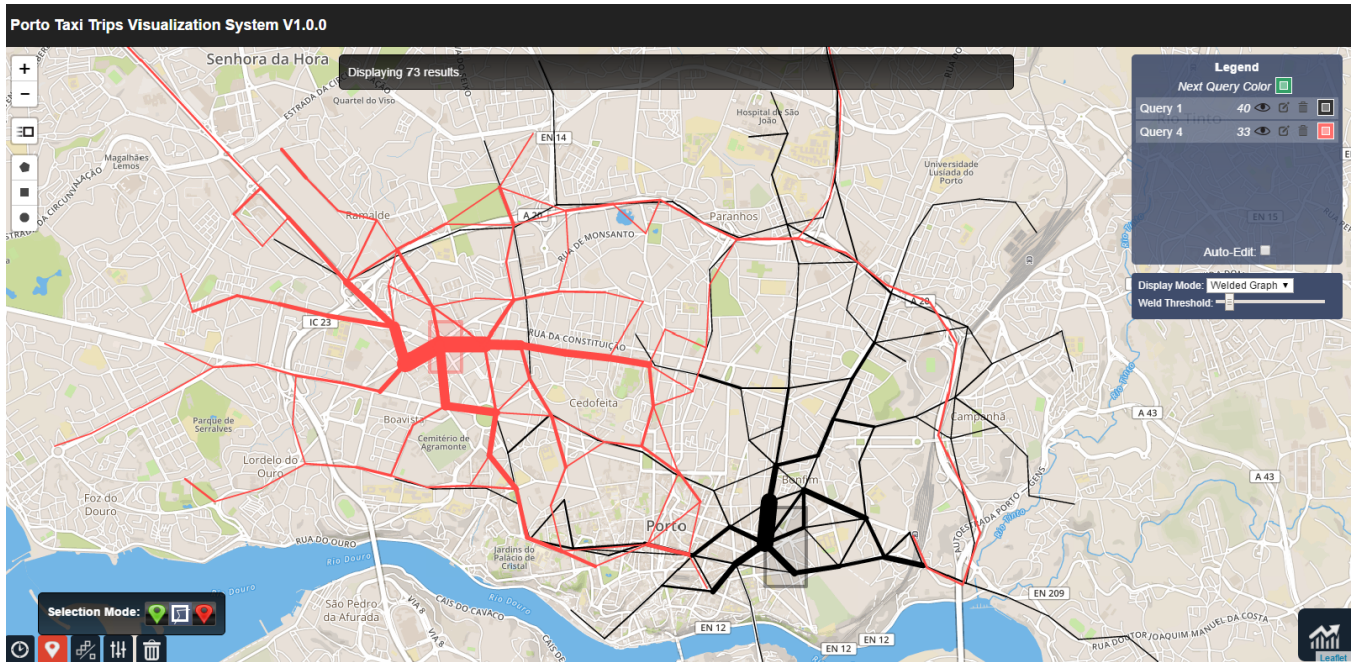


Figure 2: Interface of TrajAnalytics. Taxi trajectories in Porto from two queries are displayed with distinct colors.

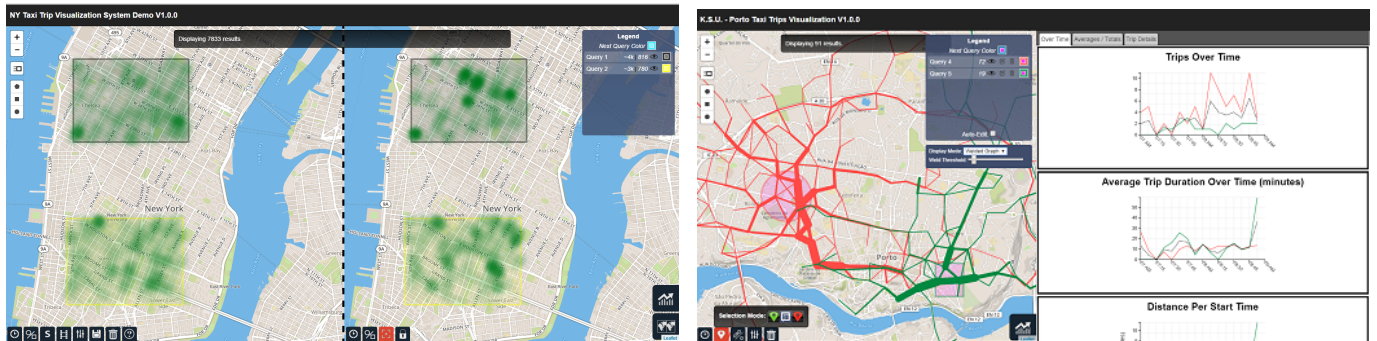


Figure 3: left: Side-by-side view to compare two queries. Right: Visual reports of the attributes of query results.

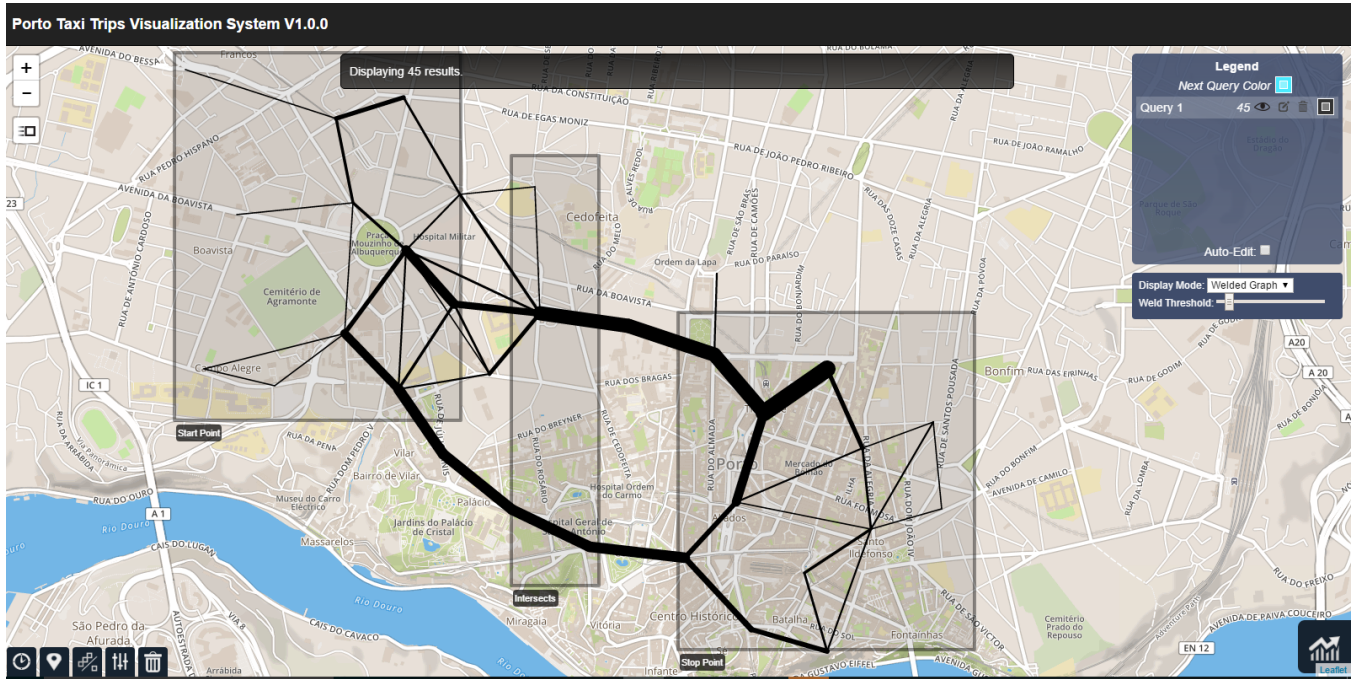


Figure 4: Interface of TrajAnalytics. Taxi trajectories in Porto are selected by joining three regional constraints.

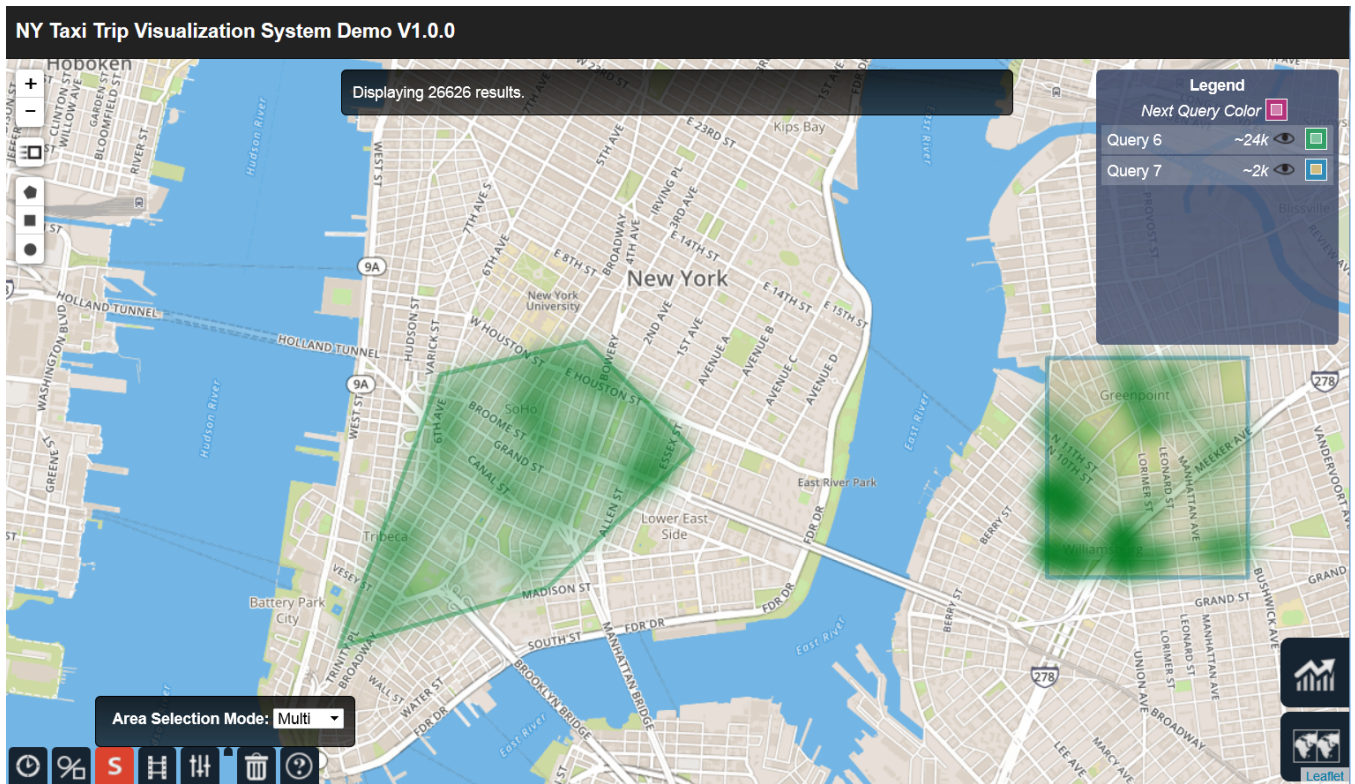


Figure 5: Interface of TrajAnalytics. Taxi pick-up locations in two select regions of New York are shown in heat maps.